

Efficiency and comparability of using new evidence platforms for updating recommendations: Experience with a type-2 diabetes guideline in Colombia

Juan Carlos Villar*¹, Luz Angela Torres López¹, Anamaría Muñoz Flórez¹, Angela Manuela Balcázar¹, Laura Parra-Gómez¹, Edgar Camilo Barrera¹

¹Research Center, Fundación Cardioinfantil-Instituto de Cardiología, Bogotá, Colombia

ABSTRACT

INTRODUCTION Updating recommendations for guidelines requires a comprehensive and efficient literature search. Although new information platforms are available for developing groups, their relative contributions to this purpose remain uncertain.

METHODS As part of a review/update of eight selected evidence-based recommendations for type 2 diabetes, we evaluated the following five literature search approaches (targeting systematic reviews, using predetermined criteria): PubMed for MEDLINE, Epistemonikos database basic search, Epistemonikos database using a structured search strategy, Living overview of evidence (LOVE) platform, and TRIP database. Three reviewers independently classified the retrieved references as definitely eligible, probably eligible, or not eligible. Those falling in the same "definitely" categories for all reviewers were labelled as "true" positives/negatives. The rest went to re-assessment and if found eligible/not eligible by consensus became "false" negatives/positives, respectively. We described the yield for each approach and computed "diagnostic accuracy" measures and agreement statistics.

RESULTS Altogether, the five approaches identified 318 to 505 references for the eight recommendations, from which reviewers considered 4.2 to 9.4% eligible after the two rounds. While PubMed outperformed the other approaches (diagnostic odds ratio 12.5 versus 2.6 to 5.3), no single search approach returned eligible references for all recommendations. Individually, searches found up to 40% of all eligible references ($n = 71$), and no combination of any three approaches could find over 80% of them. Kappa statistics for retrieval between searches were very poor (9 out of 10 paired comparisons did not surpass the chance-expected agreement).

CONCLUSION Among the information platforms assessed, PubMed appeared to be more efficient in updating this set of recommendations. However, the very poor agreement among search approaches in the reference yield demands that developing groups add information from several (probably more than three) sources for this purpose. Further research is needed to replicate our findings and enhance our understanding of how to efficiently update recommendations.

KEYWORDS Evidence-based recommendations, guidelines update, literature search, clinical practice guidelines

INTRODUCTION

Clinical practice guidelines are increasingly used to guide clinical decisions and optimize patient care [1,2]. Typically, practice guidelines include several statements with evidence-based recommendations from systematic reviews of the relevant literature. This body of evidence is the basis for

building a consensus process that incorporates stakeholders (patients, healthcare practitioners, and policy makers) with their judgments focusing on its applicability to specific situations (accessibility, economic impact, and patient preferences) [3–5]. Along with the ever-changing supporting evidence, clinical practice guidelines require regular updates to keep their recommendations up-to-date [6]. As new evidence may potentially change evidence-based recommendations, periodic updates are desirable and especially welcomed when the certainty behind the existing recommendation is low.

However, maintaining this type of document is costly and time consuming. Shekelle et al. [6], showed that most guidelines remained valid for a period of 3.6 years. The systematic review proposed by Vernooji et al. [2], included 35 methodological handbooks and showed that the most common period proposed for updating them was two to three years.

* Corresponding author jvillarc@lacardio.org

Citation Villar JC, Torres López LA, Muñoz Flórez A, Balcázar AM, Parra-Gómez L, Barrera EC. Efficiency and comparability of using new evidence platforms for updating recommendations: Experience with a type-2 diabetes guideline in Colombia. Medwave 2024;24(5):e2781 DOI 10.5867/medwave.2024.05.2781

Submitted Sep 26, 2023, **Accepted** May 16, 2024,

Published Jun 17, 2024

Postal address Fundación Cardioinfantil-Instituto de Cardiología Calle 163 A # 13B – 60 Research Center, Building H, Third Floor, Bogotá, Colombia

MAIN MESSAGES

- This is the first study to assess five different search approaches and their potential impact on the updating process of clinical practice guidelines.
- PubMed showed higher diagnostic accuracy, but the poor agreement among search approaches in the reference yield, suggests that developing groups should use several sources for updating clinical practice guidelines recommendations.
- Artificial intelligence may help expedite the updating process in the future, but there is a need for more research to refine search approaches and monitor changes in information platforms.

Although several guideline development manuals acknowledge the importance of updating clinical practice guidelines, they fail to provide specific strategies to do so. The authors also pointed to the lack of information for literature search, evidence selection and synthesis, and other aspects related to the updating process.

The Guidelines International Network provides key aspects for updating clinical practice guidelines, such as the inclusion of an expiration date in the description of the process that guideline groups will use to update recommendations [2,3]. Identifying new relevant evidence begins by building search strategies to perform a systematic reviews[4,7]. Updating approaches currently seek to incorporate new technologies and software tools to be more efficient, dynamic, and interactive [7,8]. However, the performance of these approaches remains uncertain.

In this context, we aimed to evaluate the relative contributions (in terms of yield, efficiency, and agreement) of five different search approaches to identify new relevant evidence to update evidence-based recommendations from Colombian guidelines. This assessment refers to a group of recommendations from the Colombian guidelines for type 2 diabetes mellitus (2016) [9], and the American Diabetes Association guideline (2020) [10]. This project is part of a non-communicable clinical disease practice guideline implementation project in Colombia [11].

METHODS

We selected eight recommendations from the type 2 diabetes mellitus Colombian clinical practice guidelines (2016) [9] and American Diabetes Association Guidelines (2020) [10], which included questions on prognosis or treatment (see details of the recommendations in [supplementary appendix A](#)). Our updating process used five different search approaches: (A) MEDLINE through PubMed [12], (B) The Epistemonikos database [13] using simple search, (C) The Epistemonikos database using advanced search, (D) The Living Overview of Evidence (L.OVE) platform [14] and (E) The TRIP database [15].

The authors designed search approaches based on conventional methods found in clinical practice guidelines and through the incorporation of new technological approaches ([supplementary appendices B and C](#) describe the characteristics of all search approaches and the detailed strategies used to inform each recommendation, respectively). The common goal

of our first search approach was to identify systematic reviews, potentially answering the question related to the selected recommendations. Therefore, no individual study was eligible for the update if at least one of the approaches found one or more systematic review relevant to the question.

Search strategies were generated and approved by the investigators using the same search terms (as much as possible) for all databases. In brief, we accessed MEDLINE via PubMed following a restrictive search strategy, entering the minimum number of Medical Subject Headings (MeSH) terms and keywords required, based on the search strategies published in the original guidelines, plus a narrow filter for systematic reviews identification. For the Epistemonikos (basic) search, we performed a broad search strategy using only the population and intervention terms. For the Epistemonikos advanced search, we performed an advanced search using the terms employed in the PubMed search and the filter provided by the database for the identification of systematic reviews. The L.OVE platform uses artificial intelligence to retrieve the references included in the corresponding PICO (population, intervention, control, and outcomes) question. This platform allows the addition of terms to the PICO questions that are created automatically, theoretically making search strategies more specific. We also added terms that were used in the other approaches. Finally, for the TRIP database, we performed an advanced search using pre-established terms and used the systematic reviews filter given by the database.

Study selection process

Search results were exported to RAYYAN QCRI software [16] and reviewed independently by three researchers applying predefined selection criteria. Researchers followed the elements of each PICO question to identify potentially relevant references for each search approach.

Researchers judged every document as *definitely eligible*, *eligible*, *probably eligible* or *not eligible*, falling in four possible categories. Based on their decisions, those reaching consensus (three out of three assessors agreeing) in either of the “definitely” categories were left as the final classification. The references in the “probably” categories underwent a second round of revision through at least one referee, seeking to re-classify them until a new consensus was reached. For the purpose of analysis (see below) based on diagnostic jargon, the references considered as definitely eligible/not eligible by all

assessors in the first round were labelled as “true” positives/negatives. Those undergoing re-classification in the second round and becoming definitely eligible/not eligible were labelled as “false” negatives/positives, respectively.

Data analysis

We conducted four types of data analysis. Firstly, we described the counts and proportions of references falling into each category through our classification process (i.e., retrieved references, “true positives,” “going to a second round of discussion,” etc.). Second, we formed 2×2 tables for each search approach based on the labels of diagnostic jargon and computed the usual measures of diagnostic accuracy (sensitivity, specificity, likelihood ratios, and diagnostic odds ratios), along with their 95% confidence intervals (CI). Third, we described the absolute and relative contributions (as counts and proportions) of the search approaches to finding relevant references, both individually (e.g. PubMed or L.OVE alone) and as combined approaches (e.g. included references when adding PubMed + Epistemonikos basic, or PubMed + Trip database, etc.). Finally, we computed agreement statistics across the ten possible pairs of search approaches, including the chance-corrected kappa coefficient along with 95% confidence interval. The analyses of diagnostic performance were run in Meta-DiSc, a freeware to perform meta-analyses of diagnostic studies (Clinical Biostatistics Unit of the Ramon y Cajal Research Institute (IRYCIS) from Hospital Ramón y Cajal, Madrid, Spain) [17] and all other analyses in Stata version 15 [18].

RESULTS

We updated the searches for eight recommendations using five different approaches. Four of the recommendations were taken from Colombian clinical practice guidelines [9] and four from American Diabetes Association guidelines [10]. As shown in Supplementary **Appendix A**, five of the recommendations were about treatment, two were about diagnosis of complications, and one about follow-up. Most recommendations were of moderate to high certainty of evidence, but the recommendation about follow-up had the lowest certainty of evidence (level E in the American Diabetes Association guideline).

All five approaches required a similar number of steps (see supplementary **Appendix B**), from entering the search terms to designing the search strategy to download and export the selected references. Two of the approaches exhibit some particular characteristics, such as a lack of time filter or manual selection of references, which can make the search process less efficient. For approach C, it simply found the L.OVE based on the PICO question and required less effort because it was not necessary to design and run searches. Although approach E had a search feature based on the PICO question, it did not allow the export of files, and therefore, all the strategies had to be performed using advanced features.

Retrieval process across search approaches for the eight recommendations of interest

Figure 1 shows that, despite using the same terms and keywords in the searches, the approaches varied greatly in terms of the number of articles retrieved. While approaches A, B, and E yielded a similar number of references (407, 473, and 444, respectively), approach C (Epistemonikos, advanced) returned the lowest number ($n = 318$) of references, and approach D (L.OVE) the highest ($n = 505$). As expected, most (78.7 to 90.1%) references were judged as definitely not eligible in the first round, but 8.918.2% were left for further reclassification. After the two rounds of classification, the total number of relevant references identified ranged from 20 to 31 (5.6 to 9.4% of all retrieved references). Of the 136 references considered eligible at the end of the process (some of which were repeated in two or more recommendations), two-thirds ($n = 91$) were added in the second round of discussion among the reviewers.

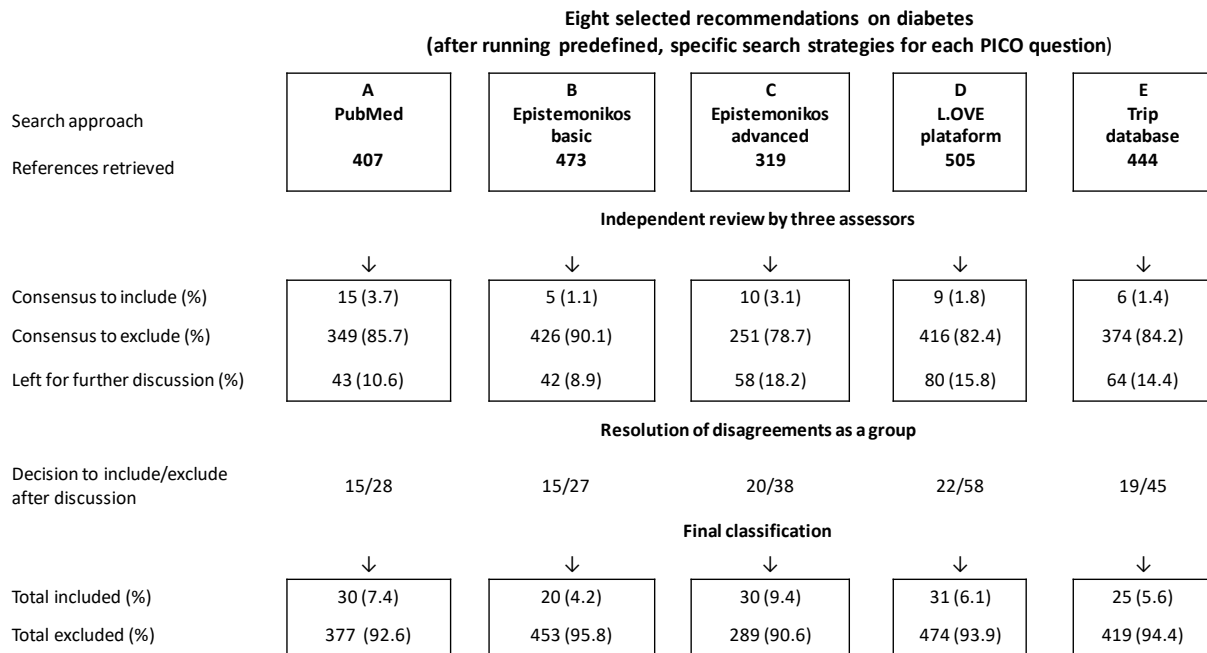
Diagnostic performance of the search approaches

Approach A (PubMed) showed the highest sensitivity (50%, while the other search approaches ranged from 24% to 33%) and the second-best specificity (93%, while the others ranged from 87% to 94%). PubMed thus outperformed the other four search approaches in terms of its positive likelihood ratio (6.73, compared with 2.23 to 4.19 of the other approaches), negative likelihood ratio (0.54, compared with 0.77 to 0.85 of the others not shown in figure) and, as a result, its diagnostic odds ratio (12.5 and 2.6 to 5.3, respectively) (Figure 2).

Distribution of the references found eligible by the reviewers

Table 1 shows the references retrieved by our five search approaches by recommendation and in total. The approaches retrieved 20 to 31 references for the eight recommendations, but Recommendation 1 accounted for 39% of all eligible references. Although PubMed showed better “diagnostic performance” among the approaches, L.OVE (D) had a higher retrieval, with 31 of the 82 documents of interest (71 single references, as some documents were eligible for more than one recommendation). None of the search approaches returned at least one reference for each of the eight recommendations of interest. Individually, these approaches contributed a relatively small portion (24% to 37%) of the eligible references. Nonetheless, the proportion of references supplied solely by each approach was relatively large (5-12 of the eligible references, i.e., 25% to 39% of their individual contributions).

Figure 3 shows the cumulative contribution of the eligible references (adding the eight recommendations) provided by the five approaches when the search sequence was started by one or the other. It is noticeable that in none of the search sequences the eligible references reached an (arbitrarily, minimal) 80% threshold ($n = 66$), even when combining any three approaches. The relative contribution of (A) PubMed appeared to be the highest among all approaches. As a result, to reach

Figure 1. Overview of the process of reference retrieval using the five search approaches.

PICO: population, intervention, control, and outcomes.

Source: Prepared by the authors according to study results.

the goal of retrieving at least 80% of the eligible references, four search approaches were necessary, and only when PubMed was included in the search.

Figure 4 shows the cumulative contribution of the searches for all recommendations starting with PubMed. Despite the variation in the yield of the recommendations, there is consistency in the need to add search approaches to reach a substantial number of eligible references. Although PubMed identified roughly half of the references for most recommendations, only in one of the eight situations (*recommendation 7*) was sufficient (to find the only relevant reference, that all other approaches also did). Moreover, in two of the recommendations (*recommendations 5 and 8*), PubMed did not return references of interest.

Agreement between search approaches

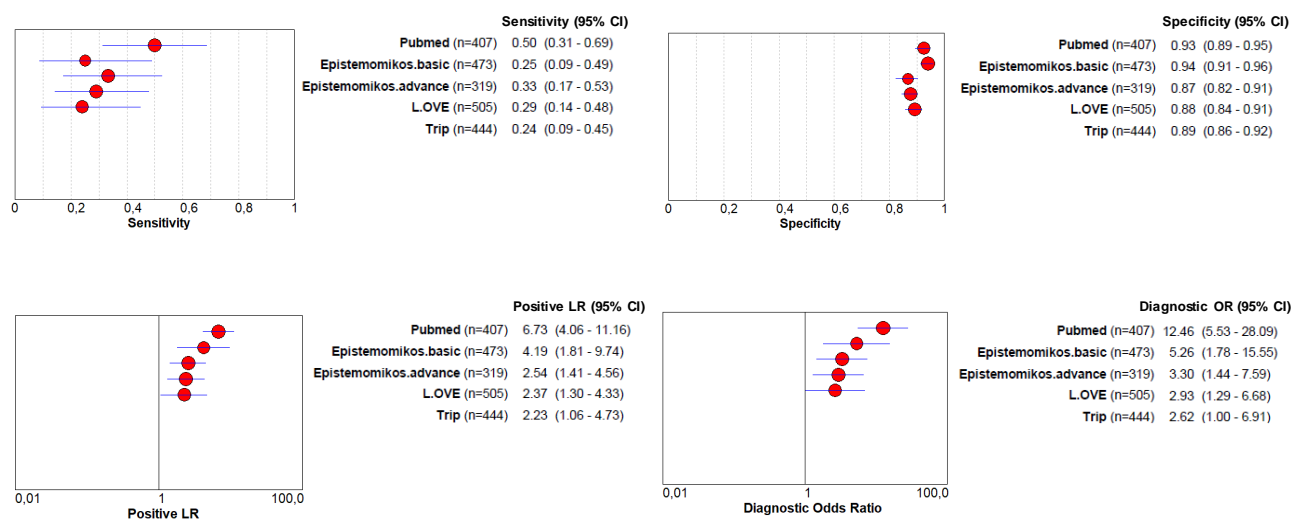
Table 2 shows the number of eligible references ($n = 71$) in agreement as found/not found (positive/negative agreement) between pairs approaches. Each cell also included the proportion of observed/expected (by chance) agreement, and the kappa statistics for the ten possible pairs of search approaches. In nine of the ten possible paired evaluations, the agreement statistics did not surpass that found by chance.

DISCUSSION

This study addressing the performance of newly available search approaches to update recommendations shows that identification of relevant studies (systematic reviews in this case) continues to be challenging. Our data indicate that feeding the review of evidence with a comprehensive set of literature requires contributions from several information platforms. We found that PubMed searches, somehow the conventional comparator, tended to be more efficient in terms of diagnostic accuracy (a higher proportion of relevant references included, with relatively fewer studies needing re-classification) compared with the other search approaches. However, despite this advantage, PubMed as a search approach (as well as others) contributed a relatively small proportion of references, and sometimes (two out of eight recommendations) did not contribute for the potential update. This is the result of a very poor agreement in the references retrieved by the search approaches under evaluation, despite a strategy to include the same (or the most similar possible) terms.

Our approach aimed primarily to identify systematic reviews in the field of type 2 diabetes, a disease that conveys a high burden and has been intensively researched. While all information platforms under assessment found potentially relevant references for the updating process, there was substantially more literature in support of treatment recommendations

Figure 2. Measures of diagnostic performance of the five approaches.



L.OVE: Living overview of evidence platform. CI, confidence interval. LR, likelihood ratio. OR, odds ratio.
Source: Prepared by the authors according to study results.

relative to other aspects. This indicates that more systematic reviews are being produced around the effect of interventions compared to other types of systematic review [19]. Regardless of the information platform, its individual contribution was modest (none over 40% of all eligible references), and no single search approach provided at least one reference for each recommendation.

In our study, the Kappa statistics between approaches proved neither clinical, nor statistical agreement. While PubMed, the more established and comprehensive search engine, had a poor agreement with the other approaches, this also seems true between the new emerging information platforms. The only approaches with agreement over chance-expected were Epistemomikos, basic search, and L.OVE platform (Cohen's Kappa 0.37, CI 95%: 0.14 to 0.60), probably because the domain of the L.OVE platform corresponds to Epistemomikos [13,14]. The poor agreement for retrieving eligible references across search approaches was also evident by showing that no combination of any three searches reached even 80% of the relevant documents. To reach such a threshold, four search approaches were necessary only when PubMed was one of them.

Conceptually, all five assessed approaches required the same steps: 1) design of the search strategy, 2) use of filters such as time or study type (systematic review), 3) selection of references, and 4) download and export files. However, those integrating the use of new technologies, such as the L.OVE platform,

might require less effort in designing the search strategy using the PICO format. In addition, in our PubMed exercise, the approach that required more elaboration for the search terms was more efficient in retrieving relevant references (diagnostic odds ratio 12.5, CI 95%: 5.5 to 28.1). The yield of references varied substantially both in general (range of 319 to 505, with a difference of more than 100 references) and for those found eligible (20 to 31 documents) among search approaches. We saw that reviewers found eligible less than 10% (range 4.2 to 9.4) of the retrieved references, as all strategies were oriented to comprehensiveness to ensure that all relevant references could be identified.

In this study we choose the consensus among three reviewers as "gold standard". The search strategies were designed by the whole research group, seeking to identify all potentially relevant references, starting from a consensus in both the terms and the studies targeted. This should have set a common ground for the reviewers and familiarity with the studies of interest for the update. The perspective of being inclusive gave the chance to all references pointed as *probably eligible* to go to a second round of review. This may have increased the number of references in need of re-assessment (8.9% to 18.2% in all search approaches), reducing efficiency, but ensuring specificity (87% to 94%) and inclusiveness (two-thirds of the eligible references came from the second round of review). This perspective guiding the selection process also explains

Table 1. Distribution of the references returned for each of the eight recommendations being updated using the five search approaches.

Diabetes recommendations under review	References included after using each search approach					Total EBR (% of each EBR within the update)
	A	B	C	D	E	
1 - Add DPP4-i to metformin if out of goals & there is no CVD	15	3	14	19	3	32 (39)
2 - Add SGLT2-i to metformin if out of goals & CVD is present	2	2	5	3	6	11 (13)
3 - Add GLP-1 (as 3 rd medication, 1 being metformin) if out of goals & BMI>30	6	1	3	5	4	14 (17)
4 - Use RAS inhibitors to prevent/reduce proteinuria	3	5	5	0	6	7 (9)
5 - Assess HbA1c goals e/3-6 months (<7% or <6.5%)	0	1	2	1	1	3 (4)
6 - Add insulin if HbA1C is >10% at diagnosis	3	6	0	3	5	12 (15)
7 - Estimate annually GFR based on serum creatinine	1	0	1	0	0	1 (1)
8 - Use monofilament test to screen for neuropathy	0	2	0	0	0	2 (2)
All recommendations (% of each approach)	30 (37)	20 (24)	30 (37)	31 (38)	25 (30)	82 *(100)
Found exclusively with this search approach (% within each contribution)	9 (30)	5 (25)	9 (30)	12 (39)	9 (36)	

DPP4-i: Dipeptidyl peptidase-4 inhibitors. EBR: Evidence-based recommendation. GLP: Glucagon-like peptide. HbA1c: Glycated hemoglobin GFR: Glomerular filtration rate. RAS: Renin-angiotensin system. SGLT2-i: Sodium-glucose cotransporter-2. BMI: Body mass index.

*There were 71 single references identified, as one may inform more than evidence-based recommendation.

Source: Prepared by the authors of this study.

the relatively low sensitivity of the searches (24% to 50%). We believe that consensus, coming from a collective assessment with no predetermined hierarchies among reviewers, may be a good approach in general, and in our case highlighted the importance of consensus in the selection process.

Our findings in context

Other authors have evaluated different approaches to identifying key references to update recommendations or to guide patient care. Garcia et al. (2015) [20] evaluated three different strategies to identify key updated references: an exhaustive approach in the literature search strategies for each clinical question, a restrictive approach using the minimum number of MeSH terms and text words, and a PLUS approach using the PLUS database. They found that the proposed method using a restrictive search strategy using some features of PubMed for MEDLINE, such as Clinical Queries filters, can be a feasible and efficient method to identify new studies that could trigger the potential update of a recommendation. In line with our results, the approach that included PubMed for MEDLINE performed well in terms of identifying relevant references and its ability to identify negative articles.

Shariff et al. (2013) [21] compared the performance of searches in PubMed and Google Scholar, aiming to find relevant references to guide patient care and not to update a specific recommendation. They found that quick clinical searches on

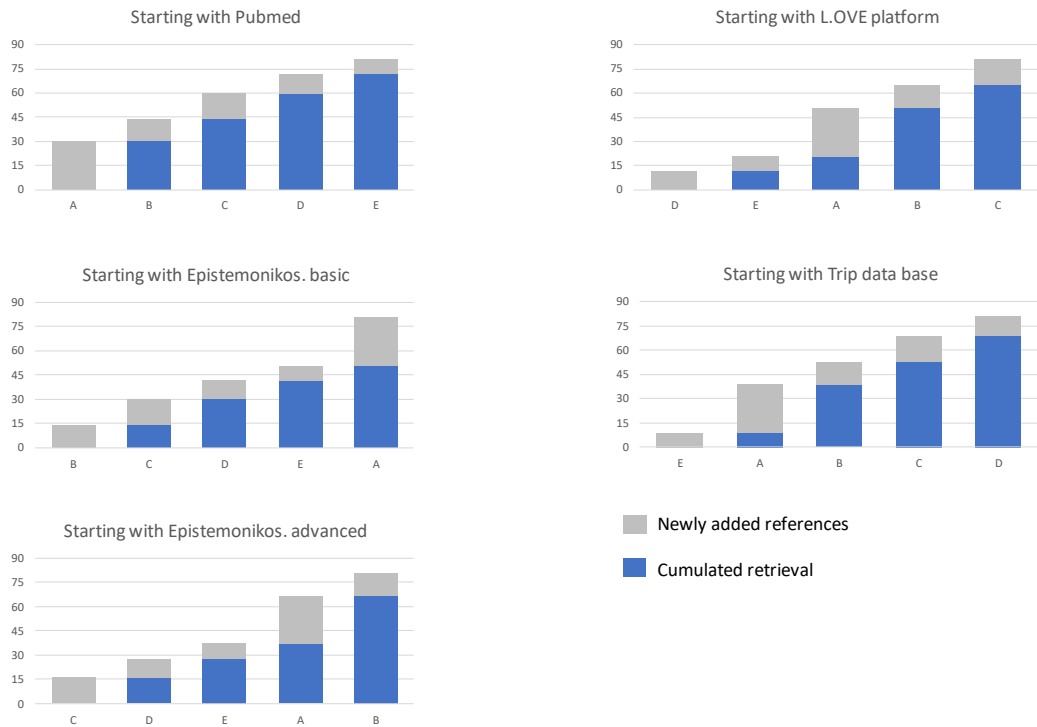
Google Scholar returned twice as many relevant articles as PubMed and provided greater access to full-text articles. We did not consider Google Scholar in any of our approaches; however, this platform resembles the use of an approach such as Epistemonikos, a basic search that does not require the design of any search strategy and still provides an important number of references [22].

Regarding combining platforms, other authors have reported that a composite approach, such as a combination of MEDLINE and Epistemonikos (complemented by reference checking of included studies), is the best combination to identify systematic review on health-related topics [23]. Rathbone et al. [24] [24] assessed seven different databases to identify relevant systematic review of interventions for hypertension and found that despite the scope of many databases, a wider search including several databases should be considered, which is also supported by our results.

Implications

In a rapidly evolving world, where new evidence may rapidly change recommendations for clinical practice, it is increasingly challenging to keep guidelines and implementation processes up-to-date. There is a growing need for efficient approaches, so recommendations change in a timely manner, as reviewing a clinical practice guideline as a whole is a resource-consuming task. The first step may be identifying which recommendations

Figure 3. Accumulative contribution of the search approaches to the eligible references according to starting approach.



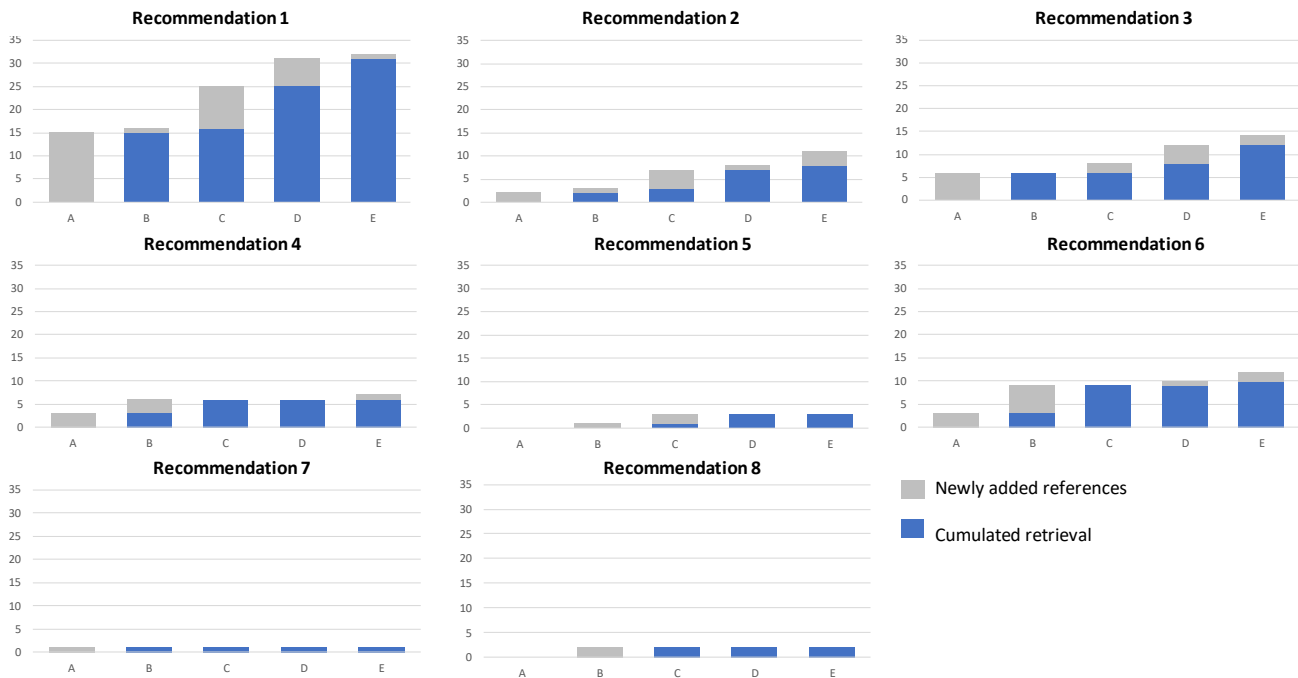
A: Pubmed, B: Epistemonikos.basic, C: Epistemonikos.advanced, D: L.OVE, E: Trip database.
Source: Prepared by the authors according to study results.

require updating, perhaps prioritizing those based on a low quality of evidence. Our findings suggest that several information platforms should be used, with PubMed as one of them. Our study also highlights the relevance of collective assessment and consensus/discussion after the first selection of articles to improve completeness in the selection process. Getting the assessors involved from the beginning, formulating the PICO questions, and co-designing the search approaches should also maintain consistent selection. Some new information platforms seem appealing and less time-consuming because they reduce the need to refine the search strategy. Other features of the databases should be considered, such as the ability to easily download and export the results and to access advanced features without a fee. Currently, these tools should be used in combination to ensure full representation in the task of updating recommendations. The vigorous wave of artificial intelligence and all its emerging possibilities should, at some point in the future, have a positive impact on reducing the time for new evidence to become an updated recommendation [25]. More research is needed to monitor the expected changes in the information platforms, so these findings, somewhat discouraging, can be refuted, and simpler search approaches can easily identify the most relevant documents for future updates.

Limitations

To the best of our knowledge, this study is the first to assess five different approaches, including those supported by artificial intelligence algorithms. The searches were designed by this team on a specific topic, targeting systematic review (rather than primary studies), and it is implied that different search strategies could lead to different results. Additionally, our searches were limited to a relatively small number of PICO questions on one topic, leading to a relatively small number of references (with low statistical precision). The validity of defining eligibility based on reviewers' criteria to find relevant (but not those finally included) references leading to real updates may be questionable and not reproducible. We think such analysis is beyond the scope of the present study.

This study did not present information about the agreement among the reviewers: the reason was an accident with the process of saving data in Rayyan. While this would have been important information (e.g., whether the three reviewers performed similarly or the need to keep all reviewers for their individual contribution to the final selection), our focus was on the performance and agreement between searches. Finally, regarding the databases evaluated, the L.OVE platform was still completing the reference classification process. Therefore, there is chance of misclassification, as several references included under each PICO set may not have been specifically related to that question.

Figure 4. Accumulative contribution of the search approaches to the eligible references using PubMed as comparator.

A: Pubmed, B: Epistemonikos.basic, C: Epistemonikos.advanced, D: LOVE, E: Trip database.

Source: Prepared by the authors according to study results.

Table 2. Agreement statistic for finding the 71 eligible references across the 10 possible pairs of search approaches.

	A PubMed	B Epistemonikos. basic	C Epistemonikos. advanced	D LOVE	E Trip database
A	\pm (% Obs-Exp) Kappa (95% CI)	8/31 (55 to 55) 0.01 (-0.22 to 0.23)	10/24 (48 to 52) -0.08 (-0.32 to 0.15)	9/25 (48 to 53) -0.10 (-0.33 to 0.13)	9/29 (54 to 54) 0.00 (-0.23 to 0.23)
B		\pm (% Obs-Exp) Kappa (95% CI)	8/30 (54 to 54) -0.01 (-0.23 to 0.21)	7/31 (54 to 55) -0.04 (-0.27 to 0.19)	12/40 (73 to 58) 0.37 (0.14 to 0.60)
C			\pm (% Obs-Exp) Kappa (95% CI)	10/25 (49 to 52) -0.06 (-0.29 to 0.17)	9/28 (52 to 53) -0.02 (-0.25 to 0.21)
D				\pm (% Obs-Exp) Kappa (95% CI)	4/25 (41 to 54) -0.29 (-0.52 to -0.06)

\pm / -, positive/negative agreement (found/not found in both search approaches). Obs-Exp, Observed-Expected (by chance) agreement. CI, Confidence interval.

CONCLUSIONS

In this study of different literature search approaches to update recommendations in type 2 diabetes no single database or combination contributed substantially to retrieve most of the eligible references. Agreement across search approaches was poor to identify relevant references. PubMed outperformed the other four search approaches in terms of efficiency and probably should be part of any update exercise, along with at least two additional databases to maintain efficiency. Further research is needed to replicate our findings in different search scenarios and enhance our understanding of how to efficiently update recommendations.

Contributor roles Study concept and design (LAT, JCV), data collection and analysis (LAT, EB, LAP, AM, AMB, JCV), drafting of the manuscript (JCV, LAT, LAP). All authors have read and approved the final manuscript.

Funding This study is part of the research project "Estudio de impacto de estrategias de información para modificar conocimientos, actitudes y prácticas en enfermedades crónicas en Bogotá" BPIN code: 2016000100037, known as "Vector Salud Bogotá" supported by local government funding (Sistema General de Regalías).

Acknowledgements We thank Maria Ximena Rojas for their input and enthusiasm in conducting this study and Lina Goetz for her support during the data collection process. We also thank Karen Moreno-Medina for helping in the editorial process of this manuscript.

Competing interests The authors declare that they have no competing interest.

Language of submission English.

Peer review and provenance Not commissioned. Externally peer-reviewed by three reviewers, double-blind.

REFERENCES

1. Eccles M, Rousseau N, Freemantle N. Updating evidence-based clinical guidelines. *J Health Serv Res Policy*. 2002;7: 98–103. <https://doi.org/10.1258/1355819021927746>
2. Vernooij RWM, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. *Implement Sci*. 2014;9: 3. <https://doi.org/10.1186/1748-5908-9-3>
3. Qaseem A, Forland F, Macbeth F, Ollenschläger G, Phillips S, van der Wees P, et al. Guidelines International Network: toward international standards for clinical practice guidelines. *Ann Intern Med*. 2012;156: 525–31. <https://doi.org/10.7326/0003-4819-156-7-201204030-00009>
4. Agbassi C, Messersmith H, McNair S, Brouwers M. Priority-based initiative for updating existing evidence-based clinical practice guidelines: the results of two iterations. *J Clin Epidemiol*. 2014;67: 1335–42. <https://doi.org/10.1016/j.jclinepi.2014.06.013>
5. Johnston ME, Brouwers MC, Browman GP. Keeping cancer guidelines current: results of a comprehensive prospective literature monitoring strategy for twenty clinical practice guidelines. *Int J Technol Assess Health Care*. 2003;19: 646–55. <https://doi.org/10.1017/s0266462303000606>
6. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? *BMJ*. 2001;323: 155–7. <https://doi.org/10.1136/bmj.323.7305.155>
7. Kredo T, Bernhardsson S, Machingaidze S, Young T, Louw Q, Ochodo E, et al. Guide to clinical practice guidelines: the current state of play. *Int J Qual Health Care*. 2016;28: 122–8. <https://doi.org/10.1093/intqhc/mzv115>
8. Vandvik PO, Brandt L, Alonso-Coello P, Treweek S, Akl EA, Kristiansen A, et al. Creating clinical practice guidelines we can trust, use, and share: a new era is imminent. *Chest*. 2013;144: 381–389. <https://doi.org/10.1378/chest.13-0746>
9. Ministerio de Salud y Protección Social, Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias). Guía de práctica clínica para el diagnóstico, tratamiento y seguimiento de la diabetes mellitus tipo 2 en la población mayor de 18 años. Bogotá D.C: Ministerio de Salud y Protección Social, Colciencias; <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/DE/CA/gpc-completa-diabetes-mellitus-tipo2-poblacion-mayor-18-anos.pdf>
10. American Diabetes Association. Introduction: Standards of Medical Care in Diabetes—2020. *Diabetes Care*. 2020. pp. S1–S2. <https://doi.org/10.2337/dc20-Sint>
11. Fundación Cardioinfantil - IC. In: Vector Salud Bogotá - LaCardio [Internet]. 2022. <https://www.lacardio.org/vector-salud-bogota/>
12. National Library of Medicine. PubMed; 2022. Available from: <https://pubmed.ncbi.nlm.nih.gov/>
13. Epistemonikos foundation. Epistemonikos; 2022. Available from: <https://www.epistemonikos.org/>
14. Epistemonikos foundation. LOVE platform; 2022. Available from: <https://iloveevidence.com/index.html>
15. Trip database Ltd. TRIP database; 2022. Available from: <https://www.tripdatabase.com/>
16. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5. <https://doi.org/10.1186/s13643-016-0384-4>
17. Zamora J, Abraira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6. <https://doi.org/10.1186/1471-2288-6-31>
18. StataCorp. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC; 2017.
19. Higgins J, Green S. Higgins J, Green S, editors. In: *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. The Cochrane Collaboration; 2011. <http://www.cochrane-handbook.org/>
20. Martínez García L, Sanabria AJ, Araya I, Lawson J, Solà I, Vernooij RWM, et al. Efficiency of pragmatic search strategies to update clinical guidelines recommendations. *BMC Med Res Methodol*. 2015;15. <https://doi.org/10.1186/s12874-015-0058-2>
21. Shariff SZ, Bejaimal SA, Sontrop JM, Iansavichus AV, Haynes RB, Weir MA, et al. Retrieving clinical evidence: a comparison of PubMed and Google Scholar for quick clinical searches. *J Med Internet Res*. 2013;15. <https://doi.org/10.2196/jmir.2624>
22. Rada G, Pérez D, Araya-Quintanilla F, Ávila C, Bravo-Soto G, Bravo-Jeria R, et al. Epistemonikos: a comprehensive database of systematic reviews for health decision-making. *BMC Med Res Methodol*. 2020;20. <https://doi.org/10.1186/s12874-020-01157-x>
23. Goossen K, Hess S, Lunny C, Pieper D. Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study. *BMC Med Res Methodol*. 2020;20. <https://doi.org/10.1186/s12874-020-00983-3>
24. Rathbone J, Carter M, Hoffmann T, Glasziou P. A comparison of the performance of seven key bibliographic databases in identifying all relevant systematic reviews of interventions for hypertension. *Syst Rev*. 2016;5. <https://doi.org/10.1186/s13643-016-0197-5>
25. Feng Y, Liang S, Zhang Y, Chen S, Wang Q, Huang T, et al. Automated medical literature screening using artificial intelligence: a systematic review and meta-analysis. *J Am Med Inform Assoc*. 2022;29: 1425–1432. <https://doi.org/10.1093/jamia/ocac066>

Eficiencia y comparabilidad del uso de nuevas plataformas de evidencia para la actualización de recomendaciones: experiencia con una guía de diabetes tipo 2 en Colombia

RESUMEN

INTRODUCCIÓN La actualización de recomendaciones de las guías de práctica clínica requiere búsquedas bibliográficas exhaustivas y eficientes. Aunque están disponibles nuevas plataformas de información para grupos desarrolladores, su contribución a este propósito sigue siendo incierta.

MÉTODOS Como parte de una revisión/actualización de 8 recomendaciones basadas en evidencia seleccionadas sobre diabetes tipo 2, evaluamos las siguientes cinco aproximaciones de búsqueda bibliográfica (dirigidas a revisiones sistemáticas, utilizando criterios predeterminados): PubMed para MEDLINE; Epistemonikos utilizando una búsqueda básica; Epistemonikos utilizando una estrategia de búsqueda estructurada; plataforma *Living Overview of Evidence* (L.OVE) y *TRIP Database*. Tres revisores clasificaron de forma independiente las referencias recuperadas como definitivamente o probablemente elegibles/no elegibles. Aquellas clasificadas en las mismas categorías "definitivas" para todos los revisores, se etiquetaron como "verdaderas" positivas/negativas. El resto se sometieron a una nueva evaluación y, si se consideraban por consenso elegibles/no elegibles, se convirtieron en "falsos" negativos/positivos, respectivamente. Describimos el rendimiento de cada aproximación, junto a sus medidas de "precisión diagnóstica" y las estadísticas de acuerdo.

RESULTADOS En conjunto, las cinco aproximaciones identificaron 318-505 referencias para las 8 recomendaciones, de las cuales los revisores consideraron elegibles el 4,2 a 9,4% tras las dos rondas. Mientras que Pubmed superó a las otras aproximaciones (odds ratio de diagnóstico 12,5 versus 2,6 a 53), ninguna aproximación de búsqueda identificó por sí misma referencias elegibles para todas las recomendaciones. Individualmente, las búsquedas identificaron hasta el 40% de todas las referencias elegibles (n=71), y ninguna combinación de cualquiera de los tres enfoques pudo identificar más del 80% de ellas. Las estadísticas Kappa para la recuperación entre búsquedas fueron muy pobres (9 de cada 10 comparaciones pareadas no superaron el acuerdo esperado por azar).

CONCLUSIONES Entre las plataformas de información evaluadas, Pubmed parece ser la más eficiente para actualizar este conjunto de recomendaciones. Sin embargo, la escasa concordancia en el rendimiento de las referencias exige que los grupos desarrolladores incorporen información de varias fuentes (probablemente más de tres) para este fin. Es necesario seguir investigando para replicar nuestros hallazgos y mejorar nuestra comprensión de cómo actualizar recomendaciones de forma eficiente.



This work is licensed under a Creative Commons Attribution 4.0 International License.